IN THE UNITED STATES PATENT AND TRADE MARK OFFICE

Applicants	:	Macquarie Research Limited; : Unisearch Limited
Serial No.	:	09/367,009
Filed	:	November 8, 1999
Title	:	Diagnosis of disease using tears
Examiner	:	M. Davis
Group Art Unit	:	1642

This is Annexure D referred to in my Declaration of

GARY COBON

GENOMICS 54, 70–78 (1998)
ARTICLE NO. GE985539

Identification of Mammaglobin B, a Novel Member of the Uteroglobin Gene Family

Robert M. Becker,*,[1] Christopher Darrow,*,[1] Drazen B. Zimonjic,† Nicholas C. Popescu,† Mark A. Watson,‡ and Timothy P. Fleming*,[2]

*Department of Ophthalmology/Box 8096 and ‡Department of Pathology/Box 8118, Washington University School of Medicine, 660 S. Euclid Avenue, St. Louis, Missouri 63110; and †Laboratory of Experimental Carcinogenesis, National Cancer Institute, National Institutes of Health, 37 Convent Drive, MSC 4255, Bethesda, Maryland 20892

Received July 10, 1998; accepted August 17, 1998

In this report, we have identified, sequenced, and characterized the expression pattern of a novel human gene, mammaglobin B. Mammaglobin B (MGB2) is highly homologous to mammaglobin (MGB1), a previously characterized human gene whose expression is limited to the mammary epithelium and frequently up-regulated in human breast cancer cells. Based upon amino acid sequence similarities, both mammaglobin and mammaglobin B may be considered members of a larger, mammalian multigene family that includes rabbit uteroglobin, human Clara Cell 10-kDa protein (CC10), and the multimeric rat prostatein protein. Together with the human CC10 gene, mammaglobin and mammaglobin B are closely linked on human chromosome 11q13. However, despite their primary sequence similarity and close chromosomal proximity, the expression of mammaglobin and mammaglobin B is nonconcordant in both nonmalignant and neoplastic tissue. © 1998 Academic Press

INTRODUCTION

Mammaglobin was identified as a breast-specific member of the uteroglobin gene family (Watson and Fleming, 1996), a multigene family with members in several mammalian species. Rabbit uteroglobin, the original member of the family, is a homodimeric secretory protein expressed primarily in the uterus, but also seen in other tissues, including the gastrointestinal tract and male urogenitary tract, and is regulated by steroid hormones (Miele *et al.*, 1994). Several properties of uteroglobin have been identified, such as its ability to serve as a substrate for transglutaminases and anti-inflammation by inhibiting phospholipase A2 (Miele *et al.*, 1994). Furthermore, a uteroglobin knockout mouse model shows that in the absence of uteroglobin, fibronectin deposits form in the kidney glomeruli, resulting in renal disease and death. Based on the transgenic model, uteroglobin is thought to prevent self-aggregation by binding to fibronectin at a higher affinity than fibronectin can self-aggregate (Zhang *et al.*, 1997a).

Rat prostatein is a tetramer composed of three subunits, C1, C2, and C3, each of which demonstrates limited but significant amino acid homology to the uteroglobin protein. The C3 peptide subunit of prostatein is sensitive to androgen regulation and has been shown to contain androgen response elements both 5′ of the transcription start site and within the first intron (Tan *et al.*, 1992). A transgenic mouse model using the 5′ regulatory region from the C3 gene driving the large T antigen (TAg) of simian virus 40 (SV40) has been shown to promote expression of TAg and subsequent tumor formation in the prostate of male mice and the mammary gland of female mice (Maroulakou *et al.*, 1994, 1997). Although prostatein is a major product of the rat prostate, no function has been defined as of yet. The genes for the C1, C2, and C3 subunits have been identified only in the rat and are either absent or poorly conserved in other mammalian species.

Members of the uteroglobin gene family identified in the human genome are limited to Clara Cell 10-kDa protein and mammaglobin. Clara Cell 10-kDa protein, the human homolog of uteroglobin, is also a homodimer and reported inhibitor of phospholipase A2 and, in rats, binds polychlorinated biphenyls (PCB) and causes accumulation of PCB in the lung (Lund *et al.*, 1985, 1988; Gillner *et al.*, 1988). Although it is expressed primarily in the lung and bronchial epithelia, CC10 is also seen in the prostate, breast, pituitary, and thyroid (Miele *et al.*, 1994). Mammaglobin, unlike uteroglobin and human CC10, is detected primarily in the breast in normal adult human tissues and appears to be overexpressed in a subset of breast tumors (Watson and Fleming, 1996). In addition to the human uteroglobin family members that have been definitively identified,

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. AF071219.

[1] Co-first authors.

[2] To whom correspondence should be addressed. Telephone: (314) 362-4981. Fax: (314) 362-3638. E-mail: fleming@am.seer.wustl.edu.

0888-7543/98 $25.00

the human prostate expresses an antigen called estramustine binding protein that has immunoreactivity to rat prostatein (Nordgren *et al.*, 1991), indicating that there are likely other members of the gene family in humans yet to be characterized. Here we describe the identification of a new uteroglobin gene family member, mammaglobin B, that shares high homology to mammaglobin, colocalizes to chromosome 11q13, and is expressed in several human tissues, including the uterus and the breast.

MATERIALS AND METHODS

Screening of phage library. Phage from a Lambda FIX II human placental genomic DNA library were plated onto agar plates with LE392 strain of *Escherichia coli*. Plaques were lifted onto Hybond-N nylon membranes (Amersham) and treated using standard methods. The probe used was derived from mammaglobin sequence from −598 to +6 relative to the transcriptional start site (Watson *et al.*, 1998) and labeled with ^{32}P using the Redi-Prime random prime labeling kit (Amersham). Hybridization at 55°C in Amersham Rapid-Hyb buffer overnight was followed by two washes in 2× SSC, 0.1% SDS at room temperature and two washes in 0.2× SSC, 0.1% SDS prior to film exposure. Hybridizing plaques were picked and replated for another two rounds of hybridization selection. The final plaque was picked and replated prior to harvest in SM (0.1 M NaCl, 0.01 M $MgSO_4$, 0.05 M Tris (pH 7.5), 0.01% gelatin). DNA was prepared with standard methods and digested with *SmaI* prior to Southern analysis. The mammaglobin 5′ probe was used to isolate hybridizing subclones, which were subsequently subcloned and sequenced using the Femtomole sequencing kit (Promega).

Chromosomal localization. Peripheral lymphocyte chromosomes were treated with methotrexate thymidine prior to fluorescence *in situ* hybridization (FISH). Mammaglobin B genomic fragment *SmaI* was labeled with biotin or digoxigenin using a random-prime DNA labeling kit (Boehringer Mannheim). FISH was performed as previously described (Zimonjic *et al.*, 1994). Slides were prepared with RNase, denatured in 2× SSC, 70% formamide for 2 min at 70°C and hybridized with 200 ng of DNA probe in 2× SSC, 50% formamide, 10% dextran sulfate, 2× Denhardt's, 1% Tween 20 for 18 h at 37°C. Posthybridization was in 0.1× SSC at 60°C and 2× SSC at 37°C for high- and low-stringency conditions. Labeled DNA was detected by fluorescein isothiocyanate avidin DCS (Vector Laboratories) or rhodamine-conjugated anti-digoxigenin (Boehringer Mannheim) as appropriate. Chromosomes were counterstained with propidium iodide or 4,6-diamino-2-phenylindole. Digital images of selected metaphase spreads were recorded during sequential excitation of each fluorochrome. Image analysis and chromosome identification were performed as previously described (Zimonjic *et al.*, 1995).

RACE PCR. Reverse transcription reactions were performed using MDA-MB 415 RNA and the Superscript II preamplification system for first-strand cDNA synthesis kit from Gibco BRL Life Technologies. Two steps of 3′ RACE PCR were done as follows: MamB RACE1 oligo (5′CTGCCACGCACGACTGAACACA) and TT7 oligo (5′TAATACGACTCACTATAGGGTTTTTTTTTTTTTTTTT) were used in PCR for 25 cycles (94°C for 2 min, 45°C for 30 s, 72°C for 45 s) followed by a 1-min extension at 72°C and 2 mM Mg^{2+}. The single band produced in this reaction (data not shown) was used as the template for a second PCR using MamB RACE2 oligo (5′CTGCACTGCTAT) and BXT7 oligo (5′GGATCCTCGAGTAATACGACTCACTATAGGG) as primers for 35 cycles (94°C for 2 min, 50°C for 1 min, 72°C for 45 s) followed by a 1-min extension at 72°C and 2 mM Mg^{2+}. The resulting product was cloned into pGem3z (Promega) and sequenced using the Femtomole sequencing kit (Promega).

Northern blots. Breast tissue samples were obtained from the Cooperative Human Tissue Network; uterus RNA samples were obtained from the Washington University Cancer Center Tumor Repository Core Facility. MDA-MB 415 RNA cells were grown to confluence in DMEM supplemented with 10% FCS, and all RNAs were prepared using the TriZol reagent from Gibco BRL Life Technologies. RNA was run on a 1% agarose, 1.5% formaldehyde gel and stained with ethidium bromide. Mammaglobin and mammaglobin B probes used for the Northern blots were full-length cDNAs labeled with the Amersham's Redi-Prime labeling kit. Blots were hybridized at 55°C overnight in Amersham's Rapid-Hyb buffer. Two 20-min washes in 2× SSC, 0.1% SDS at room temperature followed by two 45-min washes in 0.2× SSC, 0.1% SDS at 55°C were done prior to overnight exposure to film.

RNA dot blot. RNA master blot, from Clontech, was probed with the same mammaglobin B cDNA probe used to probe the Northern blots. Overnight hybridization at 60°C in Amersham's Rapid-Hyb buffer was followed by two 20-min room temperature washes in 2× SSC, 0.1% SDS and two 45-min 60°C washes in 0.2× SSC, 0.1% SDS prior to overnight exposure to film. The image containing the uterus and breast samples was digitized and formatted.

RESULTS

Identification of a New Gene Related to Mammaglobin

Screening a human genomic phage library for clones that hybridize with the 5′ end of the mammaglobin gene resulted in the identification of a clone that hybridized, but was not mammaglobin as determined by restriction analysis (data not shown). Sequence analysis of this clone demonstrated that it was a novel gene, with high homology to mammaglobin in the 150 bp 5′ of the mammaglobin transcription initiation site, as well as within exon 1 of mammaglobin. Figure 1A shows a 51% homology between the mammaglobin 5′ region and that of the hybridizing genomic clone. Based on high sequence similarity, this gene was named mammaglobin B.

To obtain a complete cDNA corresponding to the genomic clone, we targeted breast cell lines as candidates for 3′ RACE PCR. Northern analysis using a fragment of the new gene that was not homologous to mammaglobin indicated that the new gene was expressed at the mRNA level in MDA-MB 415 cells (data

FIG. 1. Mammaglobin B sequence and homologies. **(A)** Homology between mammaglobin and mammaglobin B 5′ sequences. Mammaglobin B shares 52.4% identity within the first 674 bp. Sp1 sites are in green, PEA3 sites are in orange, and AATAAATA sites are in black. **(B)** cDNA homology between mammaglobin and mammaglobin B. The cDNA sequences share a 72.8% homology, with a particularly high homology near the 5′ end of the gene. **(C)** Mammaglobin B is a member of the uteroglobin gene family of proteins, and the predicted amino acid sequence maintains conservation of residues characteristic of the family and has a higher homology to mammaglobin protein than to the other members of the family. Abbreviations: rPSC1, rPSC2, rPSC3—rat prostatein subunits 1, 2, and 3; FHG22—female hamster harderian gland cDNA; Fel-DI—feline major allergen DI; msABP-alpha—mouse salivary androgen binding protein subunit α; hMAM—human mammaglobin; hMAM-B—human mammaglobin B; hCC10—human Clara Cell 10-kDa protein; rCC10—rat Clara Cell 10-kDa protein; rUg—rabbit uteroglobin. Underlined sequence in mammaglobin B indicates the lacryglobin peptide.

A

```
Mammaglobin   : -TCAGAGAGA CCAGCAAG-- GACATGGCAG GACTGCAAGT TGCATCCAGG GTCCAAACCC TCATCCCTGA TTCCAAATTC CATCACCACC
Mammaglobin B: GGCGGAGGCT GCAGTGAGCC AAGATCGCAC CACTGCA--C TCCAGCCTGG GCAATAGAGT GAGACTCTGT CTCAAAACAA CAACAACAAC

Mammaglobin   : CACCACCCAG CACTGCCTGT CCAGCCC--- ATCAGTTACG CTACTAATTC ACAGCATCCT TTGAATCAG- -TGGAAAGAA AACCCTAAAA
Mammaglobin B: AGCAACAAAA CCCCAAAAAA CAAAAACAAA AAAACCCGCT TTTCTGAAAA AGAAATGTCT ATGGATTAGG AAGGTCTGCC TCCCATAGGA

Mammaglobin   : AGTTACCAAA GAACTAAACA AAGTTATCAA TTAATTTATA ATT-----AAC CG--AGATAA TCCTTGTAAA AA--------- TAACTTTCAG
Mammaglobin B: AGATGC--AG GTATTGCACG TAG--CCCAA ATATTTGATG ATTCCAGGAC CGGCAGATAC ACAAGTTATG AAACTGTTTG TACCTTGAAG

Mammaglobin   : CGGAGGGTGG ATGGCCTTAA ATTTCCCCAC AAAAGGTAAA ACTTTCCCAG ACATCCCTTT CTGAAAATGA GATGTCAGTG GATGAGAAAA
Mammaglobin B: AATAAGGAGA AGAGGAGGAG A--CCAAGAG GAATAGAAGA AAAAGAATAG AAGGAGGAGG AGGAAGAAGG GAAGGAGGAC TAGAAGAAAA

Mammaglobin   : GCCTGCTCCC AAGAGTAGGT GTGAGGTATT GCACCTAGGC CCACTGCTGG GATAAGCTGG CACAGATACA CGGGACATGA TGCTTTCATT
Mammaglobin B: CTACAATAAT AACAACATTG TTGAG----A CCACAAGATC ACGCCACTGG GTTTTGCT-- ---TTATATT TTGGAAA--A CACTGTCACA

Mammaglobin   : CTCACCTTAA AGAAAGAGAC AGAGAAGAAA AGGAAGGGGA GCAAGTGGAG AAGGAATACA AGAAAGAGGA AGAGAAGCAG GAAGATTCTA
Mammaglobin B: TCCATTTTAC TAAAAGCCAT GTATCATTAT TGTA----GT CCAGATTGTC AGGGTCCAAA TATTACTGCC TTCTTGGTTT CTACAAACAG

Mammaglobin   : CATACAGGCT GGCTGTGTTT CCCCTGGGGC ATGCTCCTGT TTACTGGTCC CATGCCAGGT TGACTCATTG CCTCGTTCAT GGGTGGAATT
Mammaglobin B: AGCACAGGCT GGCTGTGTTC CCATTGGTGT ACACTATTGT GTCCTGGTCC CATGCCAGGT TGACACATTG CCTCATTCCT GGGAGGGACT

Mammaglobin   : AAGATGCCTA CCTGGGGAAT AAATAGAGCA AGGCTGGGTG CTCA
Mammaglobin B: AAGGTGCCTC CCTGGGGAAT AAATAGAGCA GGGCTGGGTA CTCA
```

transcriptional start site

B

```
mammaglobin B     CCTCCACAGC AACTTCCTTG ATCCCTGCCA CGCACGACTG AACACAGACA GCAGCCGCCT
                  **********   ******** **** ***** * * ****** ***** **** ***** ****
mammaglobin       CCTCCACAGC GGCTTCCTTG ATCCTTGCCA CCCGCGACTG AACACCGACA GCAGCAGCCT

      CGCCATGAAG CTGCTGATGG TCCTCATGCT GGCGGCCCTC CTCCTGCACT GCTATGCAGA TTCTGGCTGC
      * ********  ********* ********** **********    ** ***** **** ****   *********
      CACCATGAAG TTGCTGATGG TCCTCATGCT GGCGGCCCTC TCCCAGCACT GCTACGCAGG CTCTGGCTGC

      AAACTCCTGG AGGACATGGT TGAAAAGACC ATCAATTCCG ACATATCTAT ACCTGAATAC AAAGAGCTTC
          * *** ** *  ** * *   *****  ****** *   *  * ****    ******** ***** ****
      CCCTTATTGG AGAATGTGAT TTCCAAGACA ATCAATCCAC AAGTGTCTAA GACTGAATAC AAAGAACTTC

      TTCAAGAGTT CATAGACAGT GATGCCGCTG CAGAGGCTAT GGGGAAATTC AAGCAGTGTT TCCTCAACCA
      **********  *******    ***** **  ** * ** ** **   ****  *** * **** * ** *****
      TTCAAGAGTT CATAGACGAC AATGCCACTA CAAATGCCAT AGATGAATTG AAGGAATGTT TTCTTAACCA

      GTCACATAGA ACTCTGAAAA ACTTTGGACT GATGATGCAT ACAGTGTACG ACAGCATTTG GTGTAATATG
        *  **  * *******  * *  ***   * * * *****   * * ** * ****** *    *** ** *
      AACGGATGAA ACTCTGAGCA ATGTTGAGGT GTTTATGCAA TTAATATATG ACAGCAGTCT TTGTGATTT-

      AAGAGTAATT AACTTTACCC AAGGCGTTTG GCTCAGAGGG CTACAGACTA TGGCCAGAAC TCATCTGTTG
             ** ******   * *** * **** ***** **   ** ***  ** ***  ****   ** **   *
      -----ATTTT AACTTTCTGC AAGACCTTTG GCTCACAGAA CTGCAGGGTA TGGTGAGAAA CCAACTACGG

      ATTGCTAGAA ACCACTTTTC TTTCTTGTGT TGTCTTTTTA TGTGGAAACT GCTAGACAAC TGTTGAAACC
      ******  ** *****   *   *  ** *   **********   *  ***** * ******  **********
      ATTGCTGCAA ACCACACCTT CTCTTTCTTA TGTCTTTTTA -CTACAAACT ACAAGACAAT TGTTGAAACC

      TCAAATTCAT TTCCATTTCA ATAACTAACT GCAAATC
      *   ** ***  *  **** * **** *   * *  *
      TGCTATACAT GTTTATTTTA ATAAATTGAT GGCA---        72.8% identity
```

FIG. 1—*Continued*

C

```
rPSC1          MSTVELSLCLL-I-MLAVCCYEANA---SQICELVAHETISFLMKSE--EELKKELEMYN
rPSC2          ---MRLSLCLL-T-ILVVCCYEANGQTLAGVCQALQDVTITFLLNPE--EELKRELEEFD
FHG22          ---MKLSLCLLLV-ILAVHCYEANA---ANVCPAVLSVSKSFLFKDV--EKFEAYLQTFN
Fel-DI         ---MKGARVLVLLWAALLLIWGGN----CEICPAVKRDVDLFLTGTP--DEYVEQVAQYK
msABP-alpha    ----------------------------GLCPALQRKVDLFLNGTT--EEYVQYLKEFN
rPSC3          ---MKLVFLFLLV-TIPICCYAS-----GSGCSILDEVIRGTINSTVTLHDYMKLVKPYV
hMAM           ---MKLLMVLMLA-ALSQHCYA------GSGCPLLENVISKTINPQVSKTEYKELLQEFI
hMAM-B         ---MKLLMVLMLA-ALLLHCYA------DSGCKLLEDMVEKTINSDISIPEYKELLQEFI
hCC10          ---MKLAVTLTLV-TLALCCSSAS----AEICPSFQRVIETLLMDTP--SSYEAAMELFS
rCC10          ---MKLAITITVL-MLSICCSSAS----SDICPGFLQVLEALLLGSE--SNYEAALKPFN
rUg            ---MKLAITLALV-TLALLGSPAS----AGICPRFAHVIENLLLGTP--SSYETSLKEFE

rPSC1           APPAAVEAKLEVKRC-VDQMSN-GDRLVVAETL--VYIFLKCDVKQWVETYYPEIDFYYDMN
rPSC2           APPEAVEANLKVKRC-INKIMY-GDRLSMGTSL--VFTMLKCDVKVWLQINFPRGRWFSQIN
FHG22           APPEAVKAKVEVKKC-IDSTLNYLEKMEMGKIL--AEVVGYCKGTEN
Fel-DI          ALPVVLENARILKNC-VDAKMTEEDKENALSLLDKIYTSPLC
msABP-alpha     ENRDVLDNAANIKKC-SDRTLTEEDKAQATSLINKITASRTC
rPSC3           QNHFTEKAVKQFKQCFLDQTDKTLENVGVMMEA--IFNSESCQQPS
hMAM            DDNATTNAIDELKECFLNQTDETLSNVEVFMQL--IYDSSLCDLF
hMAM-B          DSDAAAEAMGKFKQCFLNQSHRTLKNFGLMMHT--VYDSIWCNMKSN
hCC10           PDQDMREAGAQLKKL-VD-TLPQKPRESIIKLMEKIAQSSLCN
rCC10           PASDLQNAGTQLKRL-VD-TLPQETRINIVKLTEKILTSPLCEQDLRV
rUg             PDDTMKDAGMQMKKV-LD-SLPQTTRENIMKLTEKIVKSPLCM
```

Conserved Among Family Members Conserved Among Mammaglobins

FIG. 1—*Continued*

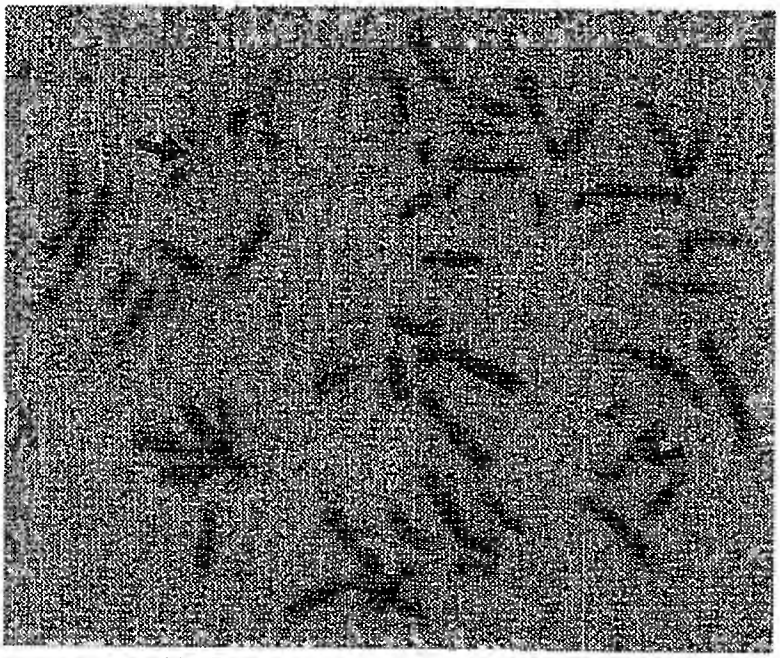

FIG. 2. FISH analysis assigns mammaglobin B to chromosome 11. Digital image of a normal human metaphase showing G-like banding after enhancement with DAPI counterstain, with both chromosomes 11 labeled at band 11q13.

not shown), a cell line that also expressed mammaglobin mRNA and protein at high levels (Watson and Fleming, 1996, unpublished observations). However, the gene was not expressed in several cell lines that did express mammaglobin, confirming the specificity of the probe and the uniqueness of the mammaglobin B gene. An oligonucleotide from the region of mammaglobin B that was highly homologous to exon 1 of mammaglobin, and hence the putative mammaglobin B exon 1, was used as a primer for 3′ RACE PCR, and a single PCR product of 511 bp was isolated and included the 3′ end of the mammaglobin B cDNA and a poly(A) tail. Sequence analysis showed a 71% homology at the DNA level to mammaglobin (Watson and Fleming, 1996). The mammaglobin B cDNA sequence (GenBank AF071219) and its homology to the mammaglobin cDNA sequence is shown in Fig. 1B. Mammaglobin B contains two introns; intron 1 is 1.7 kb in length, and intron 2 is approximately 3.3 kb in length (data not shown).

As previously reported (Watson and Fleming, 1996), mammaglobin is a member of the uteroglobin family of genes. Comparisons of the inferred protein sequence of mammaglobin B with amino acid sequences of mammaglobin and the other members of the uteroglobin gene family (Fig. 1C) show that mammaglobin B is far more similar to mammaglobin than it is to the other members of the gene family, but that it shares many of the conserved residues that characterize uteroglobin gene family members, including the cysteine residues at amino acids 22 and 99 and the lysine at amino acid 63, all of which are conserved among all family members.

Chromosomal Localization of Mammaglobin B

Mammaglobin and human CC10 both localize to 11q13 by FISH analysis (Watson *et al.*, 1998). To identify the chromosomal localization of mammaglobin B, we used a 4-kb genomic fragment of mammaglobin B sequence extending from 657 bp 5′ of the transcriptional start site through exon 1, intron 1, and exon 2 and into intron 2. Normal human chromosome spreads hybridized with biotin- or digoxigenin-labeled genomic probe had specific fluorescent signals at identical sites on both chromosomes 11 in 40 of 50 metaphases randomly selected for recordings. Twenty-five metaphases were analyzed by imaging of DAPI-generated and enhanced G-like banded chromosomes, and the fluorescence signal was localized at region 11q13 (Fig. 2). This is the same region of chromosome 11 previously shown to hybridize to both mammaglobin and human CC10. To confirm the FISH analysis, we have confirmed the presence of mammaglobin B on a genomic PAC clone that has been localized to 11q12 (data not shown).

Expression of Mammaglobin B mRNA

As mammaglobin is detected principally in the human breast epithelium and in a high percentage of human breast tumors, we wanted to examine the expression of mammaglobin B t determine if its expression pattern is similar. Figure 3 shows a human RNA

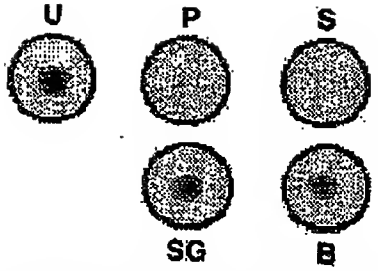

FIG. 3. RNA dot blot probed with mammaglobin B cDNA. U—uterus, P—prostate, S—stomach, SG—salivary gland, B—breast.

dot blot that contains pooled RNA samples from a collection of normal human tissues. Figure 3 shows that, like mammaglobin, mammaglobin B is expressed in the breast, but mammaglobin B is also expressed in normal uterus and salivary gland. The image shown in Fig. 3 contains the region of the dot blot that contains the breast and uterus samples; other samples on the blot did not yield a detectable signal.

Based on mammaglobin B expression in both normal breast and uterus, we examined individual RNA samples from normal and tumor breast and uterus samples. In Fig. 4, mammaglobin B exhibited a pattern of expression similar to that of mammaglobin in the breast, including expression in normal and tumor samples and apparent overexpression in tumor samples B016 and B023 in comparison to patient-matched normal samples B015 and B022. In the uterus, however, mammaglobin B expression was seen in two normal samples and in two tumor samples, while mammaglobin mRNA was seen in four of eight tumor samples and in none of the normal samples (Fig. 5). Mammaglobin B expression in the normal uterus and salivary gland shows that expression of mammaglobin and mammaglobin B is not completely concordant, as mammaglobin is never seen in the normal uterus. It is also worth noting that mammaglobin B is not expressed in the prostate, indicating that it does not appear to be a human homolog for any of the rat prostatein subunits based on its pattern of expression.

DISCUSSION

Here we have described the identification of mammaglobin B, another member of the uteroglobin gene family that is closely related to mammaglobin, but differs in its pattern of expression. We have shown that mammaglobin B is expressed in the breast, uterus, and salivary gland; Molloy *et al.* have seen the protein in tear fluid (Molloy *et al.*, 1997), and expressed sequence tags corresponding to mammaglobin B have been identified from thyroid (GenBank Accession No. AA493295), testis (GenBank Nos. AA398560 and AA393164), and ovary (GenBank No. AA525178). All of these expressed sequence tags maintain ov r 99% identity with mammaglobin B cDNA. Mammaglobin is seen only in normal breast, in a high percentage of breast tumors, and in some samples of uterine tumor tissue. Mammaglobin B expression does not appear to be detectable in every sample of a particular tissue type. There are several possible reasons for this pattern, including age and hormonal status. Mammaglobin B expression does not appear to correlate with age in the uterus samples shown in Fig. 5, and using cell culture models we have been unable to detect induction of mammaglobin B upon estrogen, progesterone, prolactin, and thyroid hormone treatment (data not shown). However, the *in vitro* milieu may not reflect the necessary conditions for gene expression changes.

Mammaglobin and mammaglobin B share very high homology in the region immediately 5′ (150 bp) of the transcription start site, yet have different patterns of expression. Mammaglobin expression is strictly limited to the breast in nontumorigenic tissue, while mammaglobin B is expressed in many normal tissues. Consequently, a study of promoter comparisons may be informative in addressing the issue of tissue specificity, particularly gene expression in the breast. This is interesting when considering the fact that both genes have been localized to 11q12–q13. The high homology and the shared locus suggest that mammaglobin and mammaglobin B arose from a duplication event from a single ancestral gene and that the divergence from this ancestral gene is solely responsible for their current differences in expression.

As stated above, mammaglobin B and mammaglobin share a very high homology in the vicinity of the transcriptional start site (Figs. 1A and 1B); based on this extremely high homology, we predict that the transcriptional start site for mammaglobin B is in the same position as reported for mammaglobin (Watson *et al.*, 1998). Neither mammaglobin or mammaglobin B has a consensus TATA box; both mammaglobin (Watson *et al.*, 1998) and mammaglobin B have a weak TATA

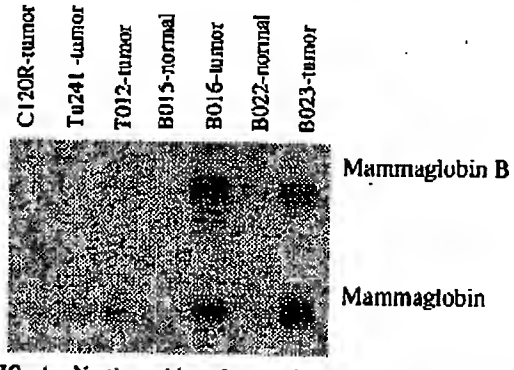

FIG. 4. Northern blot of normal and tumor breast samples probed with mammaglobin B and mammaglobin cDNAs. Samples B015 and B016 are from the same patient, as are samples B022 and B023. The same blot was probed with mammaglobin, stripped, and reprobed with mammaglobin B. Ethidium bromide staining showed comparable amounts of RNA in each lane.

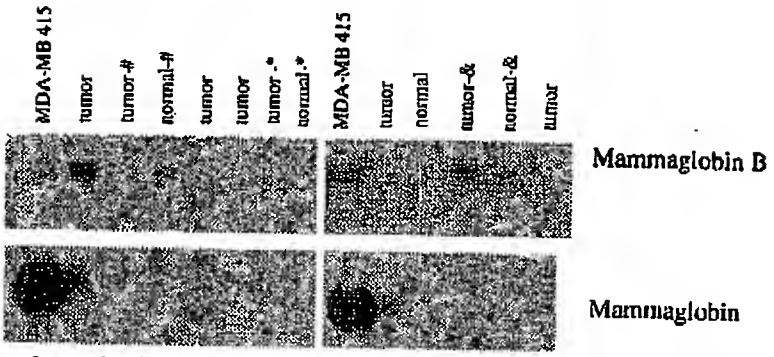

FIG. 5. Northern blots of normal and tumor uterus samples probed with mammaglobin B and mammaglobin cDNAs. Mammaglobin expression is only observed in tumor samples, while mammaglobin B expression is seen in both normal and tumor samples. The same blots were probed with mammaglobin, stripped, and reprobed with mammaglobin B. MDA-MB 415 RNA is included as a positive control on each blot. Ethidium bromide staining showed comparable amounts of RNA in each lane. #, *, and & indicate matched tumor/normal sample pairs.

sequence (5'AATAAATA) 27 bp 5' of the transcription start site (Fig. 1B). This is near the consensus region of 30 bp 5' of the transcription start site for the TATA box (Maniatis *et al.*, 1987; Dynan and Tjian, 1985; McKnight and Tjian, 1986). Like mammaglobin (Watson *et al.*, 1998), mammaglobin B promoter sequence contains polyoma enhancer-related (PEA3) motifs, but the locations of these motifs in relation to the transcription start site differ substantially from those identified in mammaglobin and are not present within repeats, as they are in mammaglobin. Mammaglobin B also contains consensus elements for Sp1 (-618, -461) binding sites (Prestridge, 1991). Although mammaglobin and mammaglobin B share some sequence motifs in the promoter region, outside of the immediate 5' 132 bp there are major differences, as would be expected in genes with different patterns of expression. Mammaglobin contains two repeated sequences within the first 1 kb of 5' sequence that have been proposed as possible regulatory sequences (Watson *et al.*, 1998). In the 651 bp of mammaglobin B 5' sequence, there are no instances of these repeat sequences or any repeats specific to mammaglobin B.

It is likely that these are not the only regulatory elements in mammaglobin and mammaglobin B, however. The C3 subunit of prostatein contains potential regulatory sequences up to 2 kb 5' of the transcriptional start site (Lund *et al.*, 1985), and progesterone receptor binding sites have been shown to exist 2.6 kb 5' of the transcription start site in rabbit uteroglobin (Jantzen *et al.*, 1987), so there is evidence for more distant regulatory elements within the uteroglobin gene family. There is as yet no simple explanation for the expression of mammaglobin B in the breast, and no evidence has shown that the putative elements seen in mammaglobin B are either necessary or sufficient for expression. Whether similar distant elements are present in mammaglobin and mammaglobin B, and their possible regulatory effects, is yet to be determined.

It has been mentioned that the amino acid sequence of the mammaglobin B protein shares conserved residues, including cysteine-22, cysteine-99, and lysine-63, that are present in every member of the uteroglobin gene family. In addition, the lysine residue at position 52, alanine-67, and serine-64 of mammaglobin B are all conserved among most family members. The complete amino acid sequence bears a 58% identity to mammaglobin, 23% to human Clara Cell 10-kDa protein, and 25% to rabbit uteroglobin. At the nucleotide level, the mammaglobin B cDNA is 71% homologous to mammaglobin (Fig. 1b), 50% to human Clara Cell 10-kDa protein, and 52% to rabbit uteroglobin. These similarities at the amino acid and nucleotide levels support our assignment of mammaglobin B as a member of the uteroglobin gene family.

There is a previously identified protein sequence that shows identity to portions of the predicted mammaglobin B gene product. In their analysis of tear proteins, Molloy *et al.* (1997) identified a peptide they call "lacryglobin" that shares homology to the mammaglobin protein. Our mammaglobin B DNA sequence predicts a protein identical to lacryglobin, with 18 additional amino acids on the N-terminus and 9 additional amino acids on the carboxy-terminus of the 68-amino-acid lacryglobin peptide. This suggests that the 18 N-terminal amino acids are cleaved prior to secretion from the lacrimal gland and that the 9 amino acids at the carboxy-terminus represent amino acids not identified in the lacryglobin fragment sequenced. The portion of the predicted mammaglobin B amino acid sequence that corresponds to the lacryglobin fragment is underlined in Fig. 1C.

We have shown here the identification of mammaglobin B, a new member of the uteroglobin gene family in humans that shares very high homology with mammaglobin but has sequence characteristics in common with all members of the gene family. Like mammaglobin and CC10, mammaglobin B 1 calizes to chromosome 11q12.3–q13.1 by FISH analysis but, unlike

mammaglobin, does not hav a highly restricted pattern of expression, being seen in the breast, uterus, salivary gland, lacrimal gland, testis, ovary, and thyroid. Mammaglobin B shares high homology to portions of the mammaglobin 5' region, but these similarities are not sufficient to confer a shared pattern of expression, making it likely that there are important regulatory elements in both genes that have yet to be discovered.

ACKNOWLEDGMENTS

This work was supported by NIH Grant CA 76227-01. Funds were provided by the Barnes/Jewish Hospital Auxiliary Fund.

REFERENCES

Dynan, W. S., and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* **316**: 774–778.

Gillner, M., Lund, J., Cembillau, C., Alexandersson, M., Hurtig, U., Bergmen, A., Klasson-Wehler, E., and Gustafsson, J. (1988). The binding of methylsulfonyl-polychloro-biphenyls to uteroglobin. *J. Steroid Biochem.* **31**: 27–33.

Jantzen, K., Fritton, H. P., Igo-Kemenes, T., Espel, E., Jantich, S., Cato, A. C., Mugele, K., and Beato, M. (1987). Partial overlapping of binding sequences for steroid hormone receptors and DNaseI hypersensitive sites in the rabbit uteroglobin gene region. *Nucleic Acids Res.* **15**: 4535–4552.

Lund, J., Brandt, I., Poellinger, A., Bergman, A., Klasson-Wehler, E., and Gustafsson, J. (1985). Target cells for the polychlorinated biphenyl metabolite 4,4'-bis(methylsulfonyl)-2,2',5,5'-tetrachlorobiphenyl. Characterization of high affinity binding in rat and mouse lung cytosol. *Mol. Pharmacol.* **27**: 314–323.

Lund, J., Nordlund, L., and Gustafsson, J. (1988). Partial purification of a binding protein for polychlorinated biphenyls from rat lung cytosol: Physicochemical and immunochemical characterization. *Biochemistry* **27**: 7895–7901.

Maniatis, T., Goodbourn, S., and Fischer, J. A. (1987). Regulation of inducible and tissue-specific gene expression. *Science* **236**: 1237–1245.

Maroulakou, I. G., Anver, M., Garrett, L., and Green, J. E. (1994). Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. *Proc. Natl. Acad. Sci. USA* **91**: 11236–11240.

Maroulakou, I. G., Shibata, M., Jorcyk, C. L., Chen, X., and Green, J. E. (1997). Reduced p53 dosage associated with mammary tumor metastases in C3(1)/TAG transgenic mice. *Mol. Carcinog.* **20**: 168–174.

McKnight, S., and Tjian, R. (1986). Transcriptional selectivity of viral genes in mammalian cells. *Cell* **46**: 795–805.

Miele, L., Cordella-Miele, E., Mantile, G., and Mukherjee, A. B. (1994). Uteroglobin and uteroglobin-like proteins: The uteroglobin family of proteins. *J. Endocrinol. Invest.* **17**: 679–692.

Molloy, M. P., Bolis, S., Herbert, B. R., Ou, K., Tyler, M. I., vanDyk, D. D., Willcox, M. D. P., Cooley, A. A., Williams, K. L., Morris, C. A., and Walsh, B. J. (1997). Establishment of the human reflex tear two-dimensional polyacrylamide gel electrophoresis reference map: New proteins of potential diagnostic value. *Electrophoresis* **18**: 2811–2815.

Nordgren, H., Nilsson, S., Eklov, S., Stattin, P., Nanning, T., Helden, T., Bergkvist, L., Kariberg, L., Lundholm, C., Harvig, B., Strang, P., Bergstrom, R., and Tribukait, B. (1991). Model DNA values and estramustine-binding protein (EMBP) as prognostic markers in prostatic cancer. *Acta Oncol.* **30**: 211–214.

Prestridge, D. S. (1991). SIGNAL SCAN: A computer program that scans DNA sequences for eukaryotic transcriptional elements. *CABIOS* **7**: 203–206.

Tan, J., Marschke, K., Ho, K., Perry, S. T., Wilson, E. M., and French, F. S. (1992). Response elements of the androgen-regulated C3 gene. *J. Biol. Chem.* **267**: 4456–4466.

Watson, M. A., and Fleming, T. P. (1994). Isolation of differentially expressed sequence tags from human breast cancer. *Cancer Res.* **54**: 4598–4602.

Watson, M. A., and Fleming, T. (1996). Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer. *Cancer Res.* **56**: 860–865.

Watson, M. A., Darrow, C., Zimonjic, D. B., Popescu, N. C., and Fleming, T. P. (1998). Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uteroglobin gene family localized to chromosome 11q13. *Oncogene* **16**: 817–824.

Zhang, Z., Kundu, G. C., Yuan, C., Ward, J. M., Lee, E. J., DeMayo, F., Westphal, H., and Mukherjee, A. B. (1997a). Severe fibronectin-deposit renal glomerular disease in mice lacking uteroglobin. *Science* **276**: 1408–1412.

Zhang, Z., Zimonjic, D. B., Popescu, N. C., Wang, N., Gerhard, D. S., Stone, E. M., Arbour, N. C., DeVries, H. G., Scheffer, H., Gerritsen, J., Colle's, J. M., TenKate, L. P., and Mukherjee, A. B. (1997b). Human uteroglobin gene: Structure, subchromosomal localization, and polymorphism. *DNA Cell Biol.* **16**: 73–83.

Zimonjic, D. B., Popescu, N. C., Matsui, T., Ito, M., and Chihara, K. (1994). Localization of the human cholecystokinin-B/gastrin receptor gene (CCKBR) to chromosome 11p15.5 → p15.4 by fluorescence in situ hybridization. *Cytogenet. Cell Genet.* **65**: 184–185.

Zimonjic, D. B., Rezanka, L., Dipaolo, J. A., and Popescu, N. C. (1995). Refined localization of the erbB-3 proto-oncogene by direct visualization of FISH signals on LUT-inverted and contrast-enhanced digital images of DAPI-banded chromosomes. *Cancer Genet. Cytogenet.* **80**: 100–102.

*605562
MAMMAGLOBIN 1; MGB1

Nucleotide, Related Entries, Protein, PubMed, LinkOut

Gene map locus 11q12.3-q13.1

TEXT

Using a modified differential display PCR technique, Watson and Fleming (1994) isolated sequence fragments that were abundantly expressed in a breast tumor relative to normal breast tissue. Using this fragment, RT-PCR, and RACE, Watson and Fleming (1996) isolated a novel full-length cDNA clone, which they designated mammaglobin-1 (MGB1). MGB1 encodes a deduced 93-amino acid protein with a 19-amino acid hydrophobic peptide signal sequence and a predicted molecular mass of 10.5-kD. The protein shares 42% sequence identity with the rat prostatein protein (rPSC3). It also shares high homology with rabbit and human uteroglobin (UGB; 192020) genes, including the conservation of cysteines known to play a role in disulfide bond formation between uteroglobin subunits and a conserved tyrosine required for progesterone binding to the uteroglobin dimer. Watson et al. (1998) reported that the arrangement and splice sites of the 3 exons of MGB1, UGB, and rPSC3 are well conserved, suggesting that these genes constitute a multigene family derived from a common ancestral sequence. The mammaglobin gene itself, however, is not well conserved phylogenetically.

Because other uteroglobin gene family members are regulated by steroid hormones, Watson et al. (1998) analyzed the MGB1 gene promoter for steroid-responsive elements. They identified 2 imperfect elements similar to estrogen- and androgen-response elements. Additionally, the promoter contains a polyoma enhancer-related (PEA3, or ETV4; 600711) motif. PEA3 is associated with activation of vimentin (193060) transcription in mammary epithelial cells and overexpression of ERBB2 (164870) in human breast tumors. Watson et al. (1998) also identified binding sites for the Sp family of transcription factors, an Ap-1 consensus site, and a binding site similar to the binding site of pregnancy-specific mammary factor in the MGB1 promoter. They did not observe any modulation of MGB1 transcription by estradiol, progesterone, dexamethasone, or androgens.

Using Northern blot analysis and RT-PCR, Watson and Fleming (1996) detected low level, breast-restricted expression of a 0.5-kb MGB1 transcript. They also detected MGB1 expression in several breast carcinoma cell lines but not in primary breast myoepithelial cells, primary breast stromal cells, an immortalized breast cell line, or an immortalized luminal ductal breast cell line. In the cell lines displaying the highest levels of expression, they also detected a 3-kb transcript, which they hypothesized was unprocessed nuclear mRNA. Watson and Fleming (1996) found that 8 of 35 primary breast carcinomas overexpressed MGB1 relative to normal breast tissue specimens, and Watson et al. (1998) found that 5 of 10 breast carcinoma cell lines and 13 of 21 metastatic breast tumors exhibited high levels of MGB1 mRNA. The overexpression did not appear to correlate with histology, tumor grade, tumor stage, or hormone receptor status. Watson and Fleming (1996) concluded that MGB1 expression is mammary-specific and may define a unique phenotype to a subset of breast carcinoma cell lines. Watson et al. (1998) presented preliminary data suggesting that expression of MGB1 is not associated with lactation, but instead with mammary gland proliferation and terminal differentiation.

By fluorescence in situ hybridization and radiation hybrid analysis, Watson et al. (1998) mapped the MGB1 gene to 11q12.3-q13.1. They detected a second FISH signal on 15q23-q24, indicating a possible homologous

gene sequence. Chromosome 11q13 is frequently amplified in breast carcinomas, but Watson et al. (1998) did not detect gene amplification or gross rearrangements in breast tumors or cell lines.

REFERENCES

1. Watson, M. A.; Darrow, C.; Zimonjic, D. B.; Popescu, N.; Fleming, T. P.:
Structure and transcriptional regulation of the human mammaglobin gene, a breast cancer associated member of the uterogl bin gene family localized to chromosome 11q13. *Oncogene* 16: 817-824, 1998.
PubMed ID: 9488047

2. Watson, M. A.; Fleming, T. P.:
Mammaglobin, a mammary-specific member of the uteroglobin gene family, is overexpressed in human breast cancer. *Cancer Res.* 56: 860-865, 1996.
PubMed ID: 8631025

3. Watson, M. A.; Fleming, T. P.:
Isolation of differentially expressed sequence tags from human breast cancer. *Cancer Res.* 54: 4598-4602, 1994.
PubMed ID: 8062249

CREATION DATE

Dawn Watkins-Chow: 1/18/2001

EDIT HISTORY

carol: 1/23/2001
carol: 1/23/2001

***604398**
MAMMAGLOBIN 2; MGB2

Nucleotide, Related Entries, Protein, PubMed, LinkOut

Alternative titles; symbols

MAMMAGLOBIN B
LACRYGLOBIN

Gene map locus 11q13

TEXT

The uteroglobin gene family includes the uteroglobin (UGB; 192020) and mammaglobin (MGB) genes. By screening a human genomic library with an MGB gene fragment, Becker et al. (1998) isolated a segment of a novel gene showing high sequence similarity to MGB. Based on this similarity, they named the novel gene mammaglobin B, which is also called mammaglobin-2 (MGB2). The authors isolated a full-length MGB2 cDNA by RACE PCR using mRNA from a human breast cancer cell line. The MGB2 gene contains 3 exons. The predicted 95-amino acid MGB2 protein contains the conserved residues characteristic of the uteroglobin family. MGB2 shares 58% amino acid sequence identity with MGB and 23% identity with UGB. Becker et al. (1998) found that MGB2 is identical to a 68-amino acid peptide of lacryglobin, a human tear protein identified and sequenced by Molloy et al. (1997), except that MGB2 contains 18 additional amino acids at the N terminus and 9 additional amino acids at the C terminus. Becker et al. (1998) suggested that the 18 N-terminal amino acids are cleaved prior to secretion from the lacrimal gland and that the 9 amino acids at the C terminus represent amino acids not identified in the lacryglobin fragment sequenced. RNA dot blot analysis found MGB2 expression in normal human breast, uterus, and salivary gland, but not in prostate or stomach. Northern blot analysis detected MGB2 expression in normal breast and breast tumor samples, with an apparent overexpression in tumor samples compared with patient-matched normal samples. Northern blot analysis of uterus samples showed MGB2 expression in 2 of 8 normal samples and 2 of 8 tumor samples.

By FISH, Becker et al. (1998) mapped the MGB2 gene to 11q13, which is where the MGB and UGB genes are located.

REFERENCES

1. Becker, R. M.; Darrow, C.; Zimonjic, D. B.; Popescu, N. C.; Watson, M. A.; Fleming, T. P. :
Identification of mammaglobin B, a novel member of the uteroglobin gene family. *Genomics* 54: 70-78, 1998.
PubMed ID : 9806831

2. Molloy, M. P.; Bolis, S.; Herbert, B. R.; Ou, K.; Tyler, M. I.; van Dyk, D. D.; Willcox, M. D.; Gooley, A. A.; Williams, K. L.; Morris, C. A.; Walsh, B. J. :
Establishment of the human reflex tear two-dimensional polyacrylamide gel electrophoresis reference map: new proteins of potential diagnostic value. *Electrophoresis* 18: 2811-2815, 1997.
PubMed ID : 9504814

CREATION DATE

Patti M. Sherman : 1/4/2000

EDIT HISTORY

mgross : 1/5/2000
psherman : 1/4/2000

Display Detailed Save Text Clip Add